Olli Tanskanen

BURST IMAGE DENOISING

Faculty of Information
Technology and Communication
Sciences
Bachelor's Thesis
April 2019

ABSTRACT

Olli Tanskanen: Burst image denoising
Bachelor's Thesis
Tampere University
Signal Processing and Machine Learning
April 2019

Modern smartphone images go through very heavy image processing before they are, for example stored, transmitted or presented on the screen. Image denoising, a process of removing noise, is one of the very first steps in smartphone image processing. Image denoising is an important step in smartphones because unprocessed images have relatively high noise levels, which can make images visually less appealing. The high noise levels are mostly caused by the relatively small sizes of the smartphone cameras and the restrictions on the exposure time. High exposure time decreases the noise levels, but also increases the blurriness of the image due to motion blur, and therefore low exposure times are often preferred.

The performance of the image denoising methods have only seen incremental improvement over the years. Also, in recent years most of the flagship smartphones have begun to use burst denoising methods. Burst denoising methods combine several images into one such that resulting image has greatly decreased noise level.

In this thesis we evaluate how burst image denoising methods compare to more traditional single-image denoising methods and what kind of future burst denoising methods hold. We also explain what the image capture process in smartphones consists of, and how both single-image and burst image denoising methods operate.

We also do experimental testing to show that burst denoising methods have a high potential in denoising performance and that they can result very good denoising results in smartphones.

Keywords: burst denoising, burst image denoising, image denoising, multi-frame denoising

CONTENTS

1. INTRODUCTION

The usage and availability of smartphones have exploded within the current decade, and with it the capturing and sharing of photos have become indisputable part of our everyday lives. With the increase of image capture, smartphone manufacturers have begun to differentiate from each other by offering better image quality year after year. The increase of image quality in smartphones is result of combination of better physical cameras and ever more sophisticated image processing methods. [1]

There are several factors that affects image quality, but one, which negatively affects image quality is *noise*. Smartphone cameras have relatively high noise levels compared for example to digital single-lens reflex DSLR) cameras due to their small sizes [2]. The small camera size restricts the available aperture size (the opening through which the light travels), and thus imposes restrictions on the amount of light that can pass inside of the camera module. Similarly, selected exposure time affects the amount of received light, higher exposure time increasing the amount of allowed light, but also increasing the *motion blur* (a blur from motion), while shorter exposure times decreases the allowed light, but also decreases the amount of motion blur.

To compensate relatively high noise levels, smartphones have complex *image denoising* methods. In recent years most flagship smartphones have begun to use burst denoising methods instead of more traditional single-image denoising methods [3,4]. Burst denoising methods take multiple images captured with short exposure time and combine them into one such that resulting image has decreased noise level due to the process of combining multiple images into one effectively corresponds to taking one image with longer exposure time while not increasing the blurriness of the image.

The purpose of this thesis it to answer to the question of why most smartphones have started to use burst denoising methods and how these methods compare to single-image denoising methods. In this thesis we take the viewpoint of real-world image capture with smartphones.

In Chapter 2 we introduce reader to smartphone camera, the processing that is done to the image and principles behind single image denoising methods. In Chapter 3 we introduce burst denoising, its advantages and disadvantages. We also describe what are the difficulties in burst denoising and how burst denoising methods works. In Chapter 4 we do experimental testing to validate the claims made in Chapter 3. Finally, in Chapter 5 we conclude the results and summarize the thesis.

2. IMAGE CAPTURE PROCESS

The process of capturing an image in smartphones is depicted in Figure 1. The image capture process begins from camera module where light is converted to digital image. The digital image is then processed by *image signal processing pipeline* where the image is processed in sequential stages with the intent of improving the final output image quality. [5] In this thesis we only address the stages marked in red as they have the most relevance of the subject of this thesis.

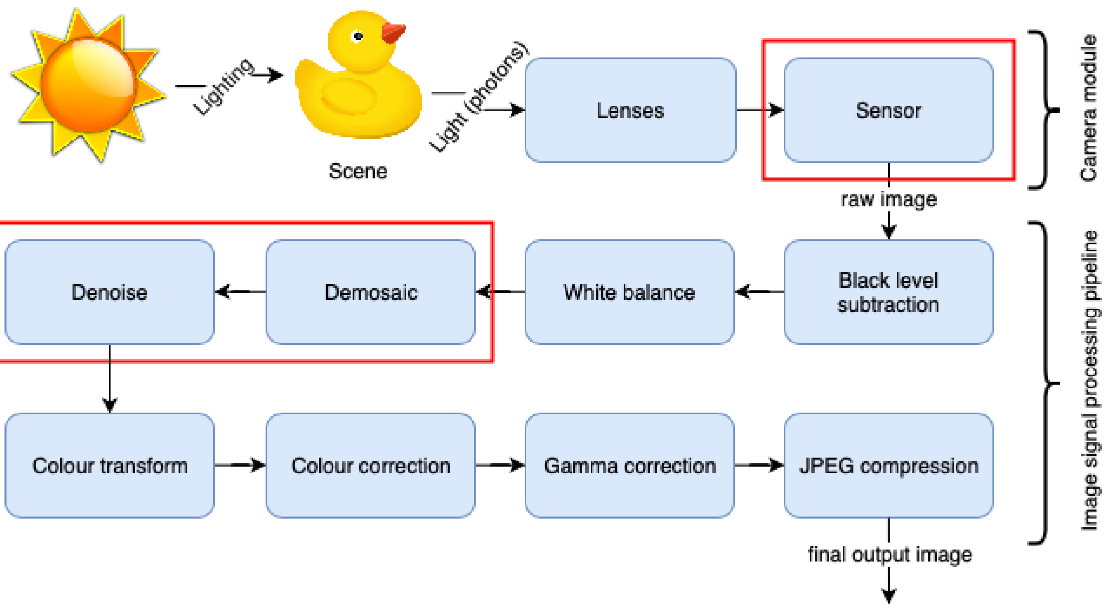

Figure 1. *An image capture process. Only stages outlined in red are addressed in this thesis.*

It should be noted that the order of stages in image signal processing pipeline can differ in real-world use case from what is stated in Figure 1. There can also be additional stages and or not all of the stages have to be implemented.

2.1 Camera module

The camera module is a modular component that include lenses, camera sensor and their housing. It can also have an optical image stabilization (OIS) solution which moves the parts of the camera module in opposite direction of the camera movement. OIS solutions can however only compensate high frequency movements, leaving most of the low frequency movement untouched. A cross section of an example camera module can be seen in Figure 2.

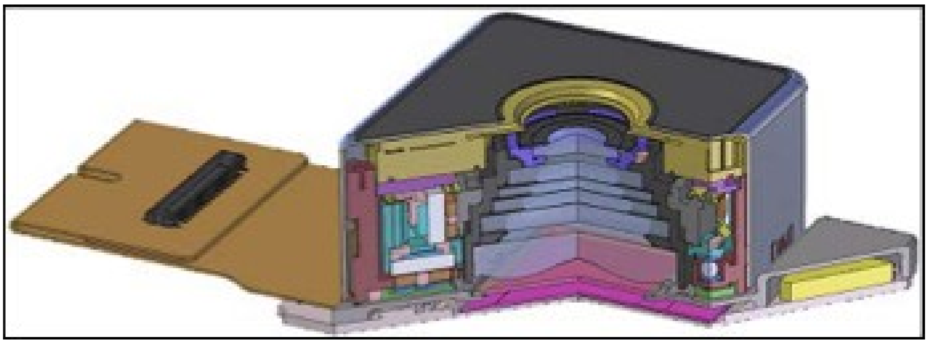

***Figure 2.** A cross-section of Microsoft Lumia 1020 camera module [5]*

The lenses, in the middle of the camera module, will collect and spread the light to the camera sensor, located at the bottom of the camera module (purple rectangle). The lenses can introduce artefacts to the image, like for example vignetting and defocus, but in this thesis they are not addressed in more detail. [2]

2.2 Camera sensor

Camera sensor will convert the light coming through the lenses to electrical charge and then into digital image.

The conversion process of converting light to electrical charge is nowadays predominantly done via complementary metal–oxide–semiconductor (CMOS) sensors. CMOS sensors have a uniform grid of photodetectors, pixels, which will accumulate electrical charge as light – photons hit them. To produce a digital image, the electrical charge of each pixel is read and then converted to digital value with analogue-to-digital converter (ADC). [6]

Due to each pixel in the sensor being sensitive to all wavelengths of light, there have to be a colour filter array (CFA) placed on top of the pixels to be able to capture a colour image. The CFA has a regular pattern of small colour filters that will filter (and let through) different wavelengths of light. This enables the sensor to capture information about all of the major colours present in the scene. Nowadays most of the camera sensors have a Bayer pattern CFA [7] (Figure 3), but different CFA patterns exists and can be used. [2]

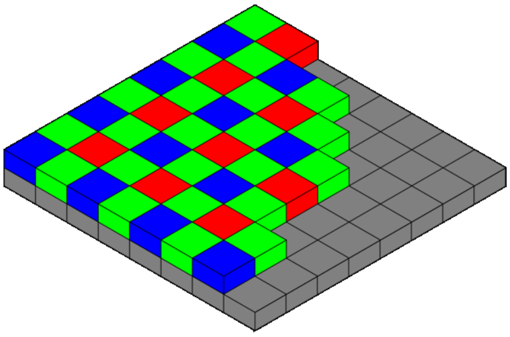

Figure 3. *A Bayer pattern CFA [8]*

The camera sensor therefore outputs a digital image where each pixel will only have information about one colour channel. This is called a *mosaiced image* or a *CFA image*. When there is no or very limited processing done to the mosaiced image, it is called *raw image*. Therefore, the output image of the camera sensor is called raw image.

The camera sensor can also add some artefacts to the raw image, for example artefacts like banding or black sun [5]. Again, in this thesis we do not go more into details of what these artefacts are or why they exist, but it is important to be aware that they exist and can affect the processing done to the image.

2.3 Noise sources & noise model

The captured raw image is corrupted by multiple noise sources. The first noise source in raw images is the process of photons hitting sensor. It is caused because the number of photons hitting the sensor fluctuates with time making it a random process [9]. This kind of noise is called *shot noise* and it can be modelled by the family of Poisson distributions. Therefore, the shot noise is signal-dependant and its variance is equal to the mean signal level, meaning that the variance is larger in brighter parts of the image and smaller in darker parts. [10]

The second noise source is thermal agitation [11], where the electrical charge in conductors will fluctuate even when there is no signal present (dark scene) producing dark current and altering the measured signal level. This kind of noise is called *thermal noise* (due to thermal agitation exponentially depending on the temperature). In low-light scenarios it can be very significant part of overall noise. [12]

The three final noise sources are *crosstalk*, *noise from gain* and quantization *noise*. The crosstalk can be modelled as a correlated noise and it is caused by the leakage of electrons from pixels to their neighbourhood pixels and leakage between readout lines of pixels. [10]

Noise from gain is caused by *analogue gain* (proportional to ISO value in camera applications), which amplifies the electrical signal level before it is converted to digital value which can further introduce noise [5].

The quantization noise is caused when the electrical signal level is converted to digital value with analogue-to-digital converter (ADC). Because all digital values are discrete, multiple analogue signal levels have to be represented with one digital value thus adding noise. The error between analogue signal level and digital value is dependent on the number of bits used to store the digital value. [6] Most camera sensors used in smartphones provide at least 10 bits of precision (1024 discrete values).

As described above, there are several noise sources that affect the raw image and due to the difficulty of separating these noise sources they are often grouped and addressed only as a one noise source. When formulating the noise model, we assume that there is a noise-free image $y \geq 0$ that the captured raw image z tries to represent but is corrupted by noise.

One popular way to characterize the captured raw image is with *Poisson-Gaussian noise* model

$$z(x) = ap(x) + n(x), \tag{1}$$

where

$$p(x) \sim \mathcal{P}(a^{-1}y(x)), \tag{2}$$

and

$$n(x) \sim \mathcal{N}(0, b), \tag{3}$$

and where $\mathcal{P}$ denotes the family of Poisson distributions, $\mathcal{N}$ denotes the Gaussian distribution and a and b are constants that depend on sensor's settings and specific characteristics both of which can be assumed to be known. The conditional mean of z is

$$E\,\{z(x) \mid y(x)\} = y(x), \tag{4}$$

and the conditional variance of z is

$$\mathrm{var}\,\{z(x) \mid y(x)\} = \sigma^2 = ay(x) + b, \tag{5}$$

where we mark the σ as the standard deviation of the noise. [10]

A signal-to-noise ratio (SNR) is a way to measure the amount of noise image has. It is defined as the signal mean level divided by the signal standard deviation, and therefore with noise model (1), the SNR is defined as

$$SNR = \frac{y(x)}{\sqrt{ay(x) + b}}, \tag{6}$$

from which we can observe that SNR can be improved by 'increasing' the noise free signal, i.e. by increasing the number of photons sensor receives. This can be controlled by increasing the aperture of the camera module, designing the camera module in a way that would decrease the a and b, or by increasing the exposure time [5]. The increase of exposure time cannot however be used in many cases as it can cause overexposure or blurriness due to motion blur. Overexposure is caused due to the fact that the camera sensor can only accumulate limited capacity of electrical charge, such that the charge will stay same even if additional photons would hit the sensor. Motion blur is caused when the camera and or scene has motion which results a blurry image. [13]

2.4 Image demosaicing

Image demosaicing or simply demosaicing is a process of converting raw image to full colour image. Image demosaicing is required due to most modern imaging software and hardware stack in smartphones can for example only store, transmit or present *full-colour images*, instead of CFA images.

Most popular demosaicing methods are *sequential* demosaicing methods, which have two distinct sequential stages of processing. The first stage in sequential demosaicing is green channel interpolation (luminance). The second stage is to interpolate the red and blue channels (chrominance) and use the interpolated green channel and the assumption that the hue is constant within an object in the image (spectral correlation). [14]

It is reasoned that this would be a good approach as the Bayer CFA has twice as many green pixels as red or blue. Green channel would therefore have more details than red or blue as it would be less likely that it would be aliased. [15] Aliasing is a phenomenon which is caused when a signal is sampled with lower sampling frequency than what it can be represented causing it to be aliased. When the aliased signal is reconstructed it will then appear as different signal to what it was originally. In images it causes moiré artefacts, where aliased high frequency details can appear as a new coarser pattern. [6]

There are several spatial methods for the interpolation of the green channel (as well as red and blue channels), most of which try to do the interpolation spatially along the edge directions as the colour homogeneity will not hold across the edge [16]. There are also methods which exploit frequency domain as it is very viable solution in cases where the spectral correlation will not hold i.e. around object boundaries [17,18].

Demosaicing methods can introduce artefacts to the image, like for example false colour, moiré, maze pattern, zipper or blur [5,16].

2.5 Image denoising

Image denoising is process of removing or attenuating noise from image. In image denoising we assume that there is a noise-free image that has been corrupted by the noise as in equation (1). Image denoising is thus a process of acquiring the noise-free image y. In this thesis we use the terms *denoising* and *image denoising* interchangeable. We also use the term *single-image denoising* which refers to the methods presented in this section.

Image denoising has been very widely researched topic in several decades and there are very wide range of different methods proposed [19]. Different methods can be categorized multiple ways but in this thesis, we categorized them as follows: *local averaging* methods, *transform thresholding* methods, *variational* methods, *patch-based* methods and *convolutional neural network* (CNN) methods [20,21].

Local averaging methods calculate intensity of each pixel by weighted average of close proximity pixels (for example within 5 by 5 area). Idea behind these methods is that close proximity pixels are likely to belong to same object or surface and therefore would have same underlying 'clean' pixel value. [22]

Transform thresholding methods decompose the image in predefined basis and based on noise statistics attenuate and/or cancel coefficients [23]. Idea is that due to the sparsity in natural images it is possible to represent a decomposed image with couple of coefficients while rest of the coefficients would represent noise. [24].

Variational methods are minimization problems, where these is an energy function E which maps the noisy image to a real number and by minimizing the function E it is possible to find the clean image. [25]

Patch-based methods finds and groups similar patches in images and combines them to reduce noise. Idea behind patch-based methods is that within natural images there exists self-similarity which can be thought to be multiple observations of the same patch [24]. Patch-based methods are good at preserving repeating textures, fine details and edges. As a disadvantage, patch-based methods can have high computational cost and artefacts in flat regions. [26,27]

CNN are methods which are trained with a set of clean and matching noisy images such that the neural network would learn how to produce a clean image from noisy one. Many of the CNN methods provide state-of-the art denoising quality on commonly used image datasets (Kodak and IMAX) but they perform relatively poorly on real world images with real world noise and can introduce additional artefacts not occurring with other methods. As an advantage, CNN methods can also have relatively low runtime. [21]

Denoising can also introduce additional artefacts to the image, like for example ringing, blur, staircase effect, checkerboard effect or blockiness [5,23]. One of the main artefacts is blurriness. In many cases a texture with small details is difficult to distinguish from noise and thus often denoising can result removal of these small details. [5] Some denoising methods can also be very aggressive in flat regions which can result an oil painted look for the image [28].

Even though single-image denoising methods are still improving year to year, most flagship smartphones have begun to use *burst denoising* methods [3,4].

3. BURST DENOISING

Burst denoising is a process where a burst of images is processed in such a way that there is only one denoised output image. In Figure 1, burst denoising would either replace only the denoising stage or both demosaicing and denoising stages depending on the burst denoising method.

Burst denoising methods are part of burst imaging methods which is a general term for methods which combine burst of images into one such a way that the output image has an improved image quality. The image quality improvements might come from high dynamic range (HDR) [13], super-resolution [29], deblurring [30] or denoising. Burst imaging process is illustrated in Figure 4.

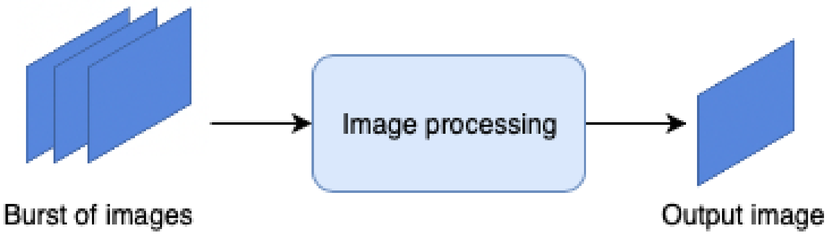

Figure 4. *The process of burst imaging*

Generally, in burst imaging or burst denoising, the burst of images can be any kind. For example, they can be raw, full colour or single channel, but in the case of burst denoising in smartphones the burst of images are raw images.

Burst denoising is also very closely related to *video denoising*. In video denoising, there is a denoised output image for each noisy input image. In the case of video denoising, the Figure 4, would have the same amount of denoised output images as input images. Video denoising methods usually work similar to burst denoising methods by combining multiple input images in the denoising process, but it is not requirement for the video denoising. Video denoising methods which work similarly to burst denoising methods can be used as a burst denoising methods by simply using only one of the denoised output image instead of all of them.

In this thesis we only concentrate in burst denoising methods. In this thesis we also use a term *frames* when we refer to a set of burst images and a term *frame* as a single image belonging to frames

3.1 Advantages

The biggest advantage of burst denoising is that it is possible to get the advantages of long exposure time without the downsides (motion blur and overexposure). In burst denoising this is mostly achieved by capturing a burst of images with short exposure time, so they are not overexposed, and they do not have any or at least very little motion blur present. The captured images are then processed such that the final image is sharp, noise-free and well exposed. [31]

As an advantage, burst denoising methods have also a big potential in denoising performance improvement. To calculate this available potential, we can think each frame to be a noisy realization of the captured scene with noise standard deviation σ. In perfect scenario where everything would be stationary and we would average the realizations, we obtain a noisy image with standard deviation of $\sigma/\sqrt{N}$, where the N, would be the number of frames (realizations) in the burst [32]. We can therefore clearly see that there is a huge potential in decreasing the variance of the noise by increasing the number of captured frames. For example, if we had noisy image with standard deviation of 40 and we captured 8 frames, in perfect scenario, by averaging the frames, we obtain an image with standard deviation of ~14 even without any additional denoising done.

3.2 Disadvantages

The biggest disadvantages in burst denoising is the increased memory and computational requirements. This can especially be problematic on devices like smartphones which have limited computational power and memory combined with the restrictions of real-time (or near real-time) image processing [31].

Most of the burst denoising methods requires that all of the frames are kept in memory while the method does the processing, increasing the memory consumption requirements manyfold due to normally the number of captured frames are from range of 2 to 15 [31,33].

Increased computational cost comes down to increase in the overall number of pixels required to process (for example a single 12-million-pixel image times 8 captured frames equals 96 million pixels). In addition, almost all of the burst denoising methods require additional stages not present in single-image denoising further increasing the computational cost.

Even though the reason burst denoising methods are used is the increase of image quality, in some cases they can result in a decrease of the image quality. Burst denoising methods can have the same artefacts as single-image denoising methods (Section 2.5). Burst denoising methods can also have additional artefacts like for example *ghosting [34].*

Figure 5. *An example of burst denoising ghosting artefact. In original frames there are two people walking from left to. Due to improper handling of moving objects, the people in input images have a 'ghost like' appearance in the burst denoised image. [34]*

Ghosting is an artefact that is caused when the burst denoising methods are unable to handle the movement of objects properly such that the output image has parts of the moving object present from each frame. [35] An example of this artefacts can be seen in Figure 5.

3.3 Difficulties

There are several difficulties in burst denoising. Images captured with smartphones are for most part taken while holding the smartphone by hand, thus inevitably introducing all kinds of movements to the camera (smartphone). This movement can be reduced if the camera module has an OIS solution. But OIS solutions are not able to solve the problem altogether as most OIS solutions are effective only in removal of high frequency movement, leaving low frequency movements untouched [36].

In addition to movement of the camera, there can be movement in the scene that is being captured. Images often include people or natural scenery, both of which can have both big and small movement happening.

Both kinds of movement can present a problem as they can cause occlusions in the captured burst of images [22]. Car driving by can occlude different parts of the road every frame, or the movement of camera can cause wall around corner become visible on one frame but not another. An example of this can be seen in Figure 6, where a car drives down the road occluding a different part of it every frame.

Figure 6. *Two (non-adjacent) frames of a burst where a car drives down the road occluding a different part of the road each frame. [34]*

In addition to occlusions, camera movement is also a problem due to parallax. Parallax is the amount of position change due to viewport (camera) movement. Objects closer to the camera have larger parallax than objects further away from the camera, causing objects to have non-linear motion. This especially is a problem if there are objects close to the camera, but less so if the captured scene is further away. [37]

There can also be changes in lighting (and thus colour) between the captured frames which can cause issues in several stages of burst denoising. This however is mostly not an issue when the number of captured frames is small as the whole capture does not last long as it is unlikely that the lighting would change during the capture. [31]

In addition to all of these problems, burst denoising methods also must to be able to handle the artefacts caused by the camera module, like vignetting or geometrical distortions [38]. Geometrical distortions caused by the lenses affect the image such that the objects closer to edges of the image are distorted more heavily, while in the middle they remain mostly unchanged. An example of geometrical distortion can be seen in Figure 7.

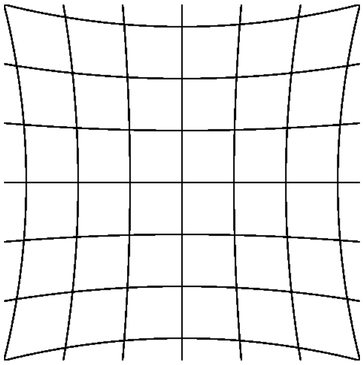

Figure 7. *An example of how geometrical distortion affects a grid of straight lines. The distortion increases with the distance from optical axes. [39]*

This presents a problem in burst denoising because an object can have a different appearance between the frames if its position in the frame changes and therefore it is distorted differently in different frames.

3.4 Stages of burst denoising

Most of the burst denoising methods can be generally split into following distinct stages: *pre-processing, image registration, image fusion* and *post-processing*. The stages of burst denoising are illustrated in Figure 8. In this thesis we do not explicitly address the post-processing stage as it is composed only from reverse operations of operations done in pre-processing stage. We however mention if the pre-processing operations require a reversing operation.

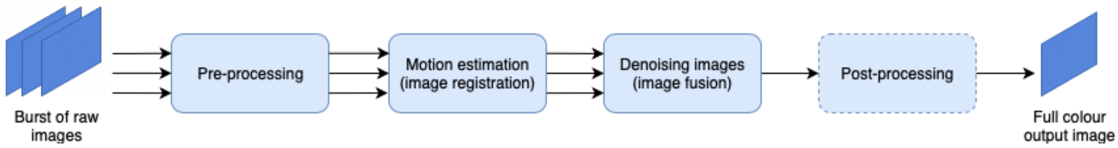

Figure 8. *The stages of burst denoising process. Three arrows mean that all of the images are carried over to next stage, whereas one arrows mean that only one image is carried to the next stage.*

It should be noted that the difference between burst denoising and other burst imaging methods comes mostly from the selected methods for the image fusion stage whereas the other stages stay mostly same.

3.4.1 Pre-processing

The pre-processing stage consists of transformations done to the frames and calculating properties of the frames. Each operation presented here is done for each frame.

Most of the methods in image fusion stage require a knowledge about the variance of the noise and thus in pre-processing stage the variance of the noise is estimated. As presented in Section 2.3, the variance can be calculated with equation (5), but due to it requiring the noise-free image y, it is not possible to directly use the equation. The noise can be estimated by for example dividing the image into patches and calculating their mean and variance [40], smoothing the image and grouping pixels with similar intensity [41] or dividing the image into patches and calculating their root-mean-square (RMS) and using the RMS value as y together with equation (5) [31].

Because the noise variance is signal-dependant, the methods have to handle per pixel variance. Alternatively, it is possible to apply a variance stabilization transform (VST) which will transform the image such that the variance of each pixel will stay constant regardless of the pixel value. This allows the methods to assume the variance to be same regardless of the pixel value. If the VST is applied, there have to be post-processing operation to reverse its effect by applying an inverse VST. [42]

The methods used in the following stages require a *reference frame*, the frame to which all other frames are compared and fused in the image registration and fusion stages. Often the reference frame is selected as the middlemost frame so that the reference frame has the least changes to all other frames [43]. Alternatively, the reference frame can be selected as the sharpest frame as some of the frames can be have blurriness due to motion blur or the frame being unfocused [31].

3.4.2 Image registration

Image registration is a process where several images from different viewpoints are transformed into one unified coordinate system by creating correspondences of pixel positions between images [38].

Image registration has been very widely studied topic along years and many different methods have been proposed. Most image registration methods used in burst denoising can be categorized either into *global homography transformation methods* or *optical flow methods.*

Global homography transform methods makes broad assumptions about the movement of the camera and the scene. They assume that there are only rotations around cameras optic centre, the objects in the 3D scene are all on same plane and the scene is far away from the camera. With these assumptions the global homography transform methods are able to greatly simplify the image registration process. As an advantage global homography transform methods have high tolerance for noise and low computational cost. As a disadvantage global homography methods cannot handle parallax. [22,37,44]

In global homography transform methods, the homography transformation is usually obtained by feature matching, like for example SIFT [45]. These work by finding distinctive features from image and matching them across all of the frames, from which the movement of each feature can be extracted and the homography calculated.

There are also methods which finds the homography in frequency domain. These kinds of methods usually add additional restrictions to the assumptions, like that any movement is parallel to camera plane. Frequency domain-based methods exploit the fact that planar rotation and shift in spatial domain are phase changes in frequency domain thus simplifying the problem. Frequency domain methods are simple, resilient to noise and have very low computational cost compared to other methods. [46]

The second category of image registration methods is **optical flow methods**. Optical flow methods calculate motion at every pixel location by computing spatial and temporal derivatives. Optical flow methods have high accuracy but many of them have extremely high computational cost which makes then unsuitable for smartphones. [38] As a disadvantage they are also not robust to noise and will often misregister movements in high noise scenarios. They can also have difficulties with occlusions [34].

Because both kind of methods have their advantages and disadvantages, methods used in smartphones are usually a combination of several different methods with high accuracy traded for lower computational cost and better noise resilience. [31]

3.4.3 Image fusion

An image fusion is a process of fusing several images into one such a way that the resulting image would have less noise when compared to any single image. A good image fusion method would be both robust to misalignments caused by the image registration and it would be able to handle occlusions well [47].

Many of the image fusion methods used in burst denoising are single-image denoising methods extended to work with set of burst images. This is especially simple in local averaging methods or in patch-based methods where one can just extend the search neighbourhood to set of burst images. The neighbourhood extension usually increases the computational cost of the method significantly and also these kinds of methods do not handle occlusions that well [31,34].

There are also methods developed explicitly for image fusion. These kinds of methods usually still use same constructs as single-image denoising methods, but they are used explicitly in temporal context. One example of this kind of method is temporal averaging, which calculates the average of the pixel value along the frames after motion compensation. Usually these methods also have additional denoising done spatially. [48]

It should be noted that the image fusion method can also do demosaicing while doing the fusion, but it requires a sub-pixel precision from the image registration and interpolation in image fusion which increases the computational cost. If the image fusion method does not do the demosaicing, normal demosaicing methods (Section 2.4) can be used after the burst denoising [31].

4. EXPERIMENTAL RESULTS

To evaluate the claims that burst denoising methods have a big potential in denoising performance and to see how well current burst denoising methods perform in smartphones, we did two kinds of experimental testing. First, we tested that burst denoising methods are suitable in the context of smartphone image capture by taking everyday pictures with smartphone. Secondly, we evaluated different burst denoising methods numerically in synthetic tests.

4.1 Real world tests

Real world tests were done using OnePlus 5T with Google Camera application (ported to OnePlus 5T [49]). The Google Camera application has a *HDR*+ [50] mode. When the HDR+ mode is turned off, the application will only take one image while when the HDR+ mode is turned on it will take a burst of images and combine them to output one image (i.e. burst denoising). Images were taken both with HDR+ mode turned off and with HDR+ mode turned on.

We captured images in several scenarios; daylight outdoor, good indoor lighting and bad indoor lighting. In all of the scenarios we let the Google Camera application to automatically choose the proper exposure time and ISO settings. The smartphone was also held in hand while capturing the images. The number of captured frames was set to be 12. We also took several images of the scene and selected ones that were most representative of the average image quality among the images.

Because there is no ground truth image, all of the comparisons are done visually. Because we turn the HDR+ mode on and off between the shots, it should be noted that the camera might have moved between the shots and thus images can have small differences in content and colour constancy.

(a) (b)

Figure 9. *Examples of artefacts that can appear in our burst denoised images due to the Google Camera application is not properly ported to work with OnePlus 5T camera sensor. (a) a dot pattern artefact, no dots should be present in the image (b) lens shading artefact, image should have completely uniform brightness across the image.*

Also, because to the Google Camera application is not officially supported on OnePlus 5T device, the captured images can have some artefacts present due to the application port was not properly adapted to work with the OnePlus 5T camera sensor. These artefacts are not present in Google Camera's officially supported devices and are not caused by the burst denoising method. Therefore, the artefacts seen in Figure 9 should not be taken into account in the image quality comparison.

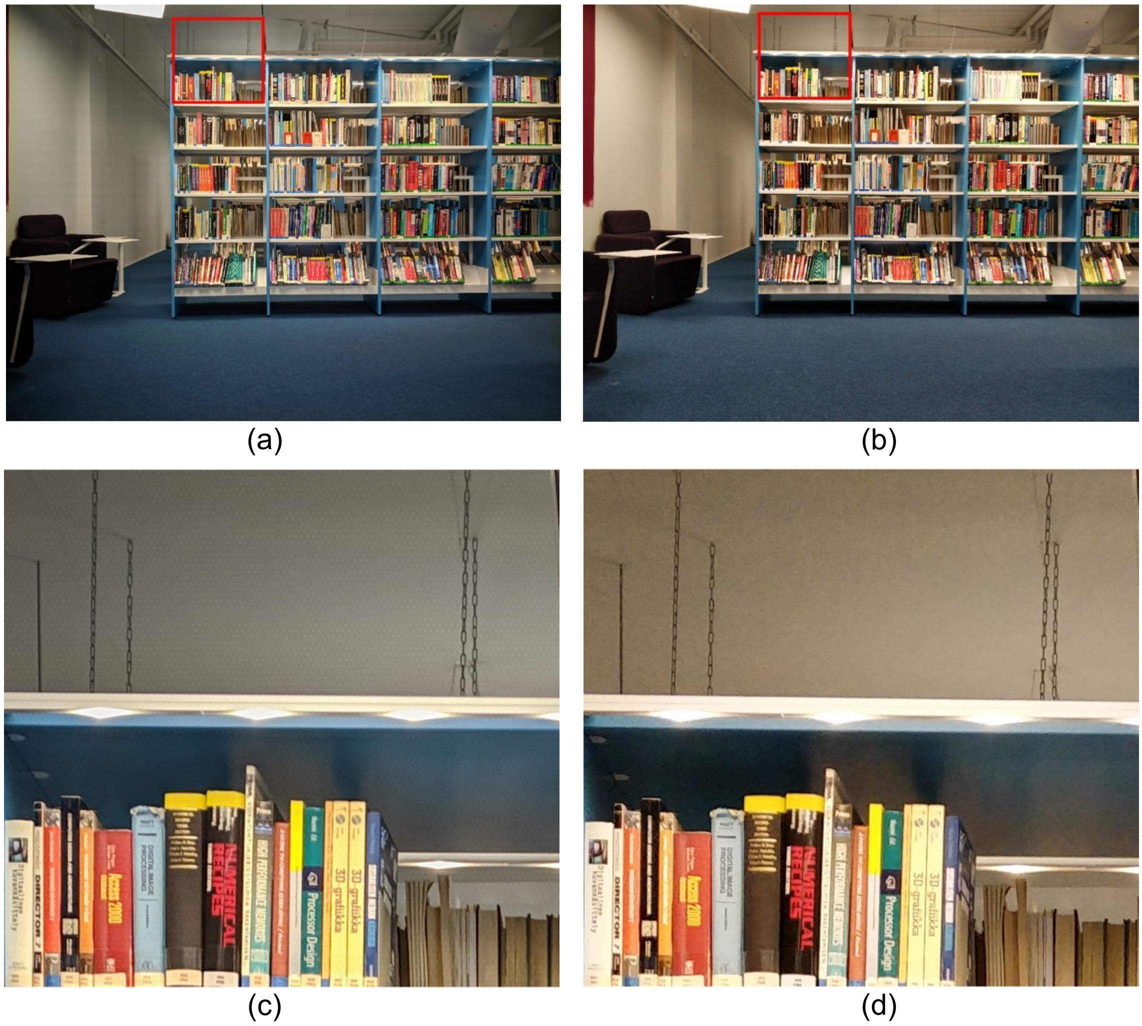

Figure 10. *Images captured in good lighting conditions. (a) full HDR+ on image (b) full HDR+ off image (c) cropped HDR+ on image (d) cropped HDR+ off image*

In Figure 10 we can see that there are some differences in image quality between the single-image and burst denoised images. In Figure 10 (c) we can see that the ceiling is noise-free, and all of the chains are very clearly defined, whereas in Figure 10 (d) the ceiling has very visible noise and some of the chains are a bit unclear.

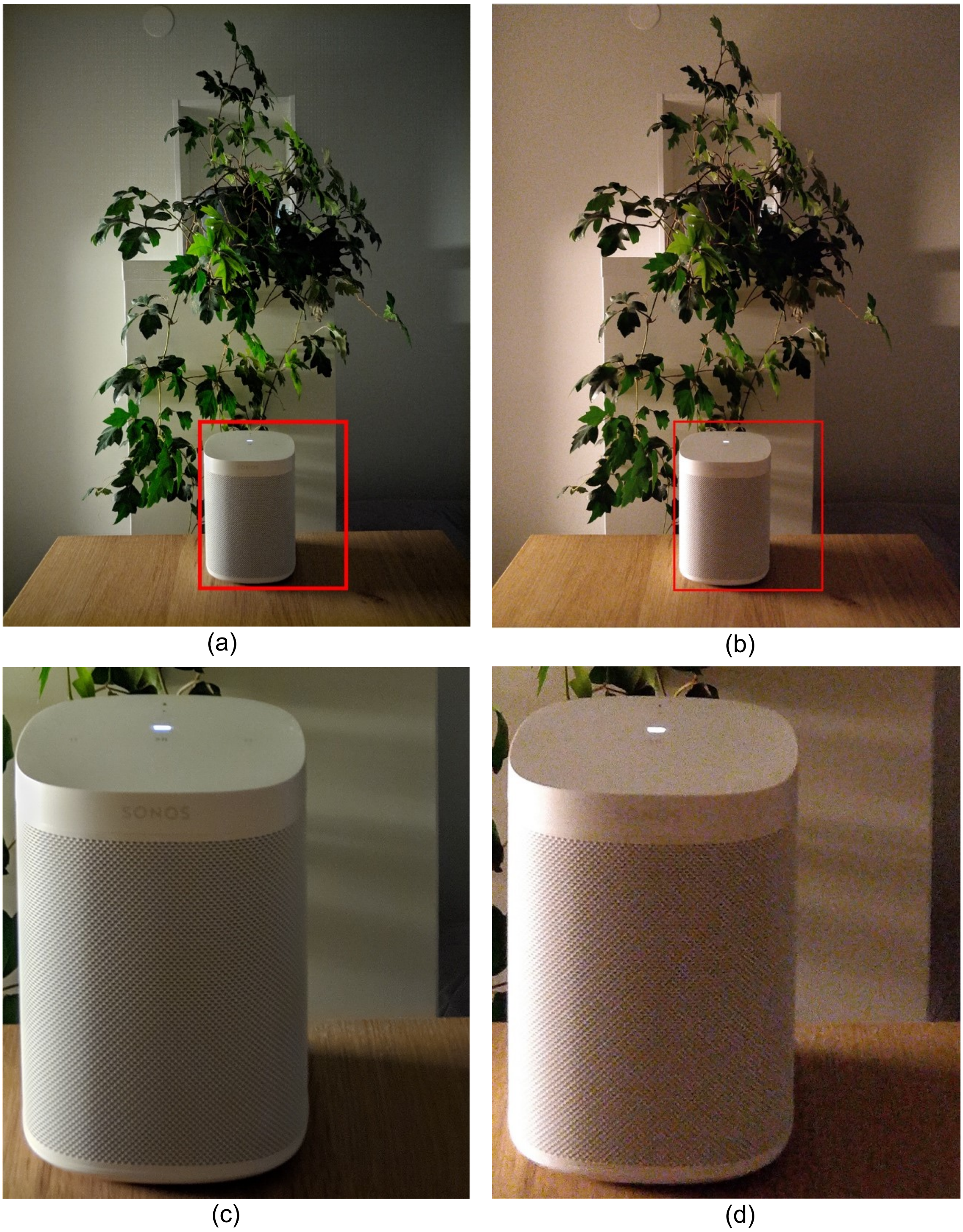

Figure 11. *Images captured in bad lighting conditions. (a) full HDR+ on image (b) full HDR+ off image (c) cropped HDR+ on image (d) cropped HDR+ off image*

In Figure 11 we can see that the burst denoised image has considerably better image quality than the single-image denoised image. Figure 11 (d) has visible noise in all areas, even on flat regions, while in Figure 11 (c) there is mostly perceivable noise in the flat regions.

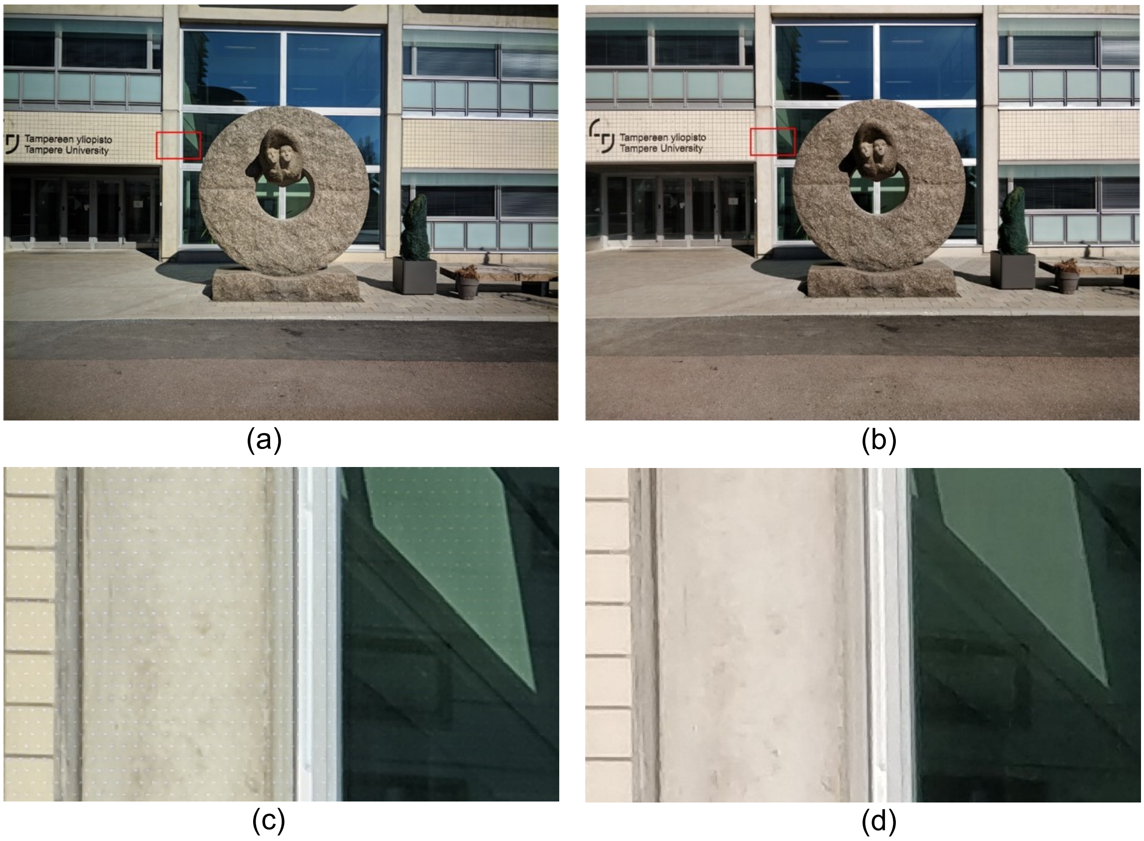

Figure 12. *Images captured in sunny outside conditions. (a) full HDR+ on image (b) full HDR+ off image (c) cropped HDR+ on image (d) cropped HDR+ off image*

In Figure 12 we can see that in sunny outside conditions there is no big differences between the single-image and burst denoised images. Biggest image quality difference between the two can be seen in the structures of the image. In Figure 12 (d) there is fuzziness in the structures whereas in Figure 12 (c) all structures are very clearly defined.

4.2 Synthetic tests

Synthetic tests were done to get a numerical value for the image quality of denoised images. With synthetic tests we are able to see qualitatively not only how burst denoising methods compare to single-image denoising methods, but also how burst denoising methods perform with different noise levels and number of frames used in denoising. We are also able to see how close burst denoising methods are to the best-case scenario.

We used state-of-the-art burst and single-image denoising methods. For single-image denoising methods we selected C-BM3D [51] and NL-Bayes [52]. For burst denoising we chose SPTWO [22] and V-BM4D [53]. Both of the burst denoising methods are *video denoising* methods, but can be used as a burst denoised methods as stated in Section

3. Implementations for all of the methods were provided by their respective authors [54–56]. Even though both of the burst denoising methods were video denoising methods, they can still be used as burst denoising methods by taking only one of the outputted frames as the denoised result.

The comparisons were done by denoising the same noisy image sequence with every selected method and comparing the denoised results to clean reference image by using peak-signal-to-noise ratio (PSNR) as a performance metrics.

For input image sequences, we used popular burst image dataset Middlebury [57]. The dataset provides clean, noise-free full-colour image sequences. The dataset has mostly natural scenes with people or objects moving.

Because the Middlebury dataset consists of noise-free image sequences, we added additive white Gaussian noise (AWGN) to the images. For colour images, the noise was added separately for each colour channel.

The Poisson-Gaussian noise model, presented in Section 2.3, cannot be used with full colour images as it is a noise model for raw images. Also, because the noise model for raw images will change very complex ways in the demosaicing step, in our synthetic tests we use the AWGN model for the sake of simplicity.

All of the methods were provided with the real variance of the AWGN model. All of the methods also used only default parameters provided by their authors.

The number of frames burst denoising methods used in their denoising process is told case by case. It should be noted that the number of frames used has no effect single-image denoising methods which only use one frame (reference frame) in their denoising process.

Even though this test setup is not directly comparable to denoising of raw images captured with smartphones [58], it can however show how burst denoising methods behave with different noise levels or with different number of frames. We also argue that the relative denoising performance order between different denoising methods do show a trend that is applicable to the case of denoising real-world images captured with smartphones due to results of our real-world tests and the results shown in [58] which have only a small changes in image quality between the methods when the denoising is done in raw or full-colour images.

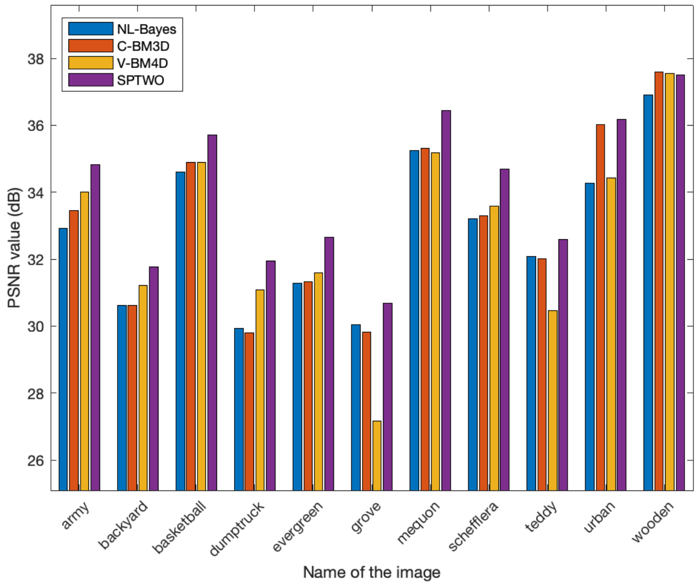

Figure 13. *PSNR values of each method of burst sequences from Middlebury dataset. Both burst denoising methods used maximum number of frames. ‘grove’ and ‘teddy’ have 2 frames while other sequences 8 frames. The standard deviation of the noise is* $\sigma = 20$*.*

In Figure 13 we can see that in all cases burst denoising methods produce better PSNR values. We can also see that in cases where the is only 2 frames, V-BM4D perform worse than both single-image denoising methods, but when there are more frames available it usually beats them.

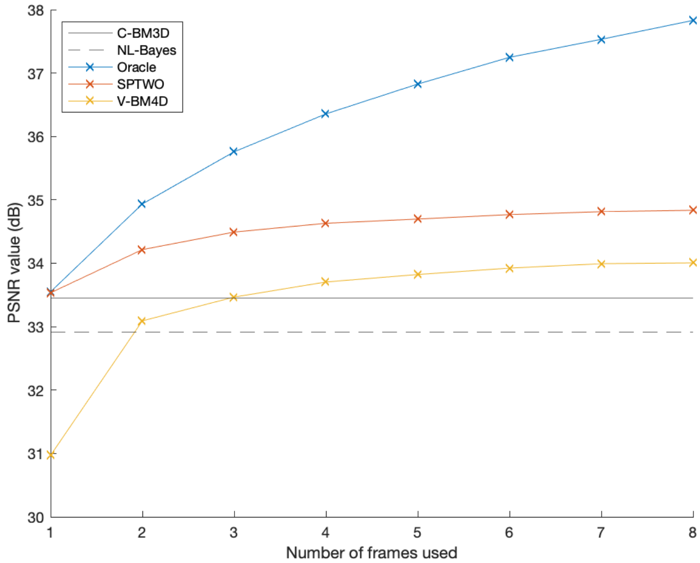

Figure 14. *PSNR values of denoised images with respect to the number of frames used in denoising. Denoised image sequence was ‘army’ with standard deviation of* $\sigma = 20$*.*

The ‘Oracle’ in Figure 14 is implemented by taking the first frame of the sequence and taking new realizations of the noisy image. The realizations were then averaged and resulting image was used as an input to the SPTWO method. Number of frames in figure depicts the number of realizations used in the averaging and therefore, the SPTWO method is used as a single-image denoising method (only 1 frame used in denoising) in the case of ‘Oracle’. The ‘Oracle’ therefore depicts PSNR values which could be achieved in the best-case scenario when using the SPTWO method.

In Figure 14 we can see that V-BM4D has a substantial increase in PSNR values when the number of frames starts to increase but starts to even out when the number of frames is higher than 5. We can also see similar behaviour with SPTWO method though the overall increase in PSNR values is not as big as with the V-BM4D. We can also see that there is a big difference between the ‘Oracle’ and SPTWO method and thus there is a big available potential in the future development of burst denoising methods.

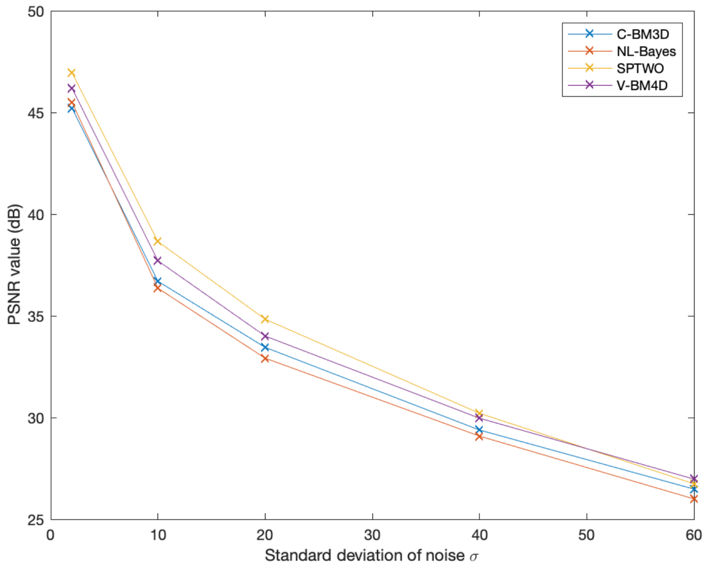

Figure 15. *PSNR values of denoised images with respect to the standard deviation of noise. Used image sequence is ‘army’ with 8 frames.*

In Figure 15 we can see that the image quality decreases relatively same rate with all of the methods. This is important to know as some burst denoising methods can perform poorly with higher noise levels as the image registration methods can produce incorrect motion estimation under high noise levels [34].

Figure 16. *A single noise-free frame of the ‘army’ image sequence*

Figure 17. *A single frame with added AWGN with standard deviation of deviation $\sigma = 20$.*

Figure 18. *The denoised output of C-BM3D method. PSNR value for the denoised output is 33.45 dB.*

Figure 19. *The denoised output of the SPTWO method with all 8 input frames. PSNR value for the denoised output is 34.84 dB.*

In Figure 19 we can see that overall SPTWO produces visually pleasing results, but it is unable to restore the finest details seen in the bottommost closeup in Figure 16. The structures in Figure 19 are also clear and sharp and the image does not have any created artefacts for example in the yellow cardboard box.

In Figure 18 we can see that the denoised output of C-BM3D method has decreased image quality compared to the denoised output of SPTWO in Figure 19. Overall Figure 18 is more blurred and it has some fuzziness seen in topmost closeup and in the surface of the yellow cardboard box.

5. CONCLUSIONS

The purpose of this thesis was to answer to the question of why most smartphones have started to use burst denoising methods and how well these methods compare to single-image denoising methods. In this thesis we have shown that in real-world image capture with smartphones burst denoising methods produce an order of magnitude better denoising results compared to single-image denoising methods. We have also shown that even current state-of-the-art burst denoising methods have a huge potential in denoising performance improvements and therefore it is self-evident that why most flagship smartphones have begun to use burst denoising methods.

As burst denoising becomes more common and new solutions come to overcome its difficulties it is also expected that burst denoising methods will also spread from flagship smartphones to more affordable ones. We also expect new methods to reach more of the available denoising performance potential and thus see even better results in the coming years.

Overall this thesis is a good starting point for understanding modern image capture process of smartphones while also giving a deeper look into the future methods used in image processing and demonstrating their possible potential and effectiveness.

Future work could take into account the whole image processing pipeline and how artefacts and errors caused in the early stages of the pipeline affect the final image quality. It could also go deeper into burst denoising and especially research what are the failure cases of burst denoising methods and in what point current methods ‘break down’ and what could be done to make them more robust.

REFERENCES

[1] Breakthrough Mobile Imaging Experiences: How Qualcomm Snapdragon processors are transforming photography [Internet]. [cited 2019 Apr 17]. Available from: https://www.qualcomm.com/documents/whitepaper-breakthrough-mobile-imaging-experiences

[2] Eismann K, Duggan S, Grey T. Real world digital photography. 2nd ed. Berkeley (CA): Peachpit Press; 2004.

[3] If Portrait Mode is So Great, Why Did Samsung and Huawei Use DSLRs? [Internet]. The Phoblographer. 2018 [cited 2019 Apr 18]. Available from: https://www.thephoblographer.com/2018/12/16/if-portrait-mode-is-so-great-why-have-samsung-and-huawei-used-dslr-photos/

[4] Shankland S. Brace yourself, because digital photography will change more in the next 10 years than in the last 10 [Internet]. CNET. [cited 2019 Apr 18]. Available from: https://www.cnet.com/news/digital-photography-gets-radical-changes-smartphone-mirrorless-tech/

[5] Peltoketo V-T. Benchmarking of Mobile Phone Cameras. 2016.

[6] Jähne B. Practical handbook on image processing for scientific and technical applications, second edition. 2004.

[7] Bayer BE. Color imaging array. US3971065A, 1976.

[8] Bayer pattern on sensor [Internet]. [cited 2019 Apr 18]. Available from: https://commons.wikimedia.org/wiki/File:Bayer_pattern_on_sensor.svg

[9] Mandel L. Fluctuations of Photon Beams: The Distribution of the Photo-Electrons. Proceedings of the Physical Society. 1959;

[10] Azzari L, Borges L, Foi A. Modeling and Estimation of Signal-Dependent and Correlated Noise. In: Denoising of Photographic Images and Video: Fundamentals, Open Challenges and New Trends. 2018. p. 1–36.

[11] B. Johnson J. Thermal Agitation of Electricity in Conductors. APS Journals. 1928;

[12] Wang X. Noise in sub-micron CMOS image sensors. 2008.

[13] Reinhard E, Heidrich W, Debevec P, Pattanaik S, Ward G, Myszkowski K. High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting. 2 edition. Burlington, MA: Morgan Kaufmann; 2010.

[14] li X, Gunturk B, Zhang L. Image demosaicing: A systematic survey. Proc SPIE 6822. 2008;

[15] Gunturk BK, Altunbasak Y, Mersereau RM. Color plane interpolation using alternating projections. IEEE Transactions on Image Processing. 2002;

[16] K. Gunturk B, Glotzbach J, Altunbasak Y, W. Schafer R, M. Mersereau R. Demosaicking: Color filter array interpolation. Signal Processing Magazine, IEEE. 2005;

[17] Losson O, Macaire L, Yang Y. Comparison of Color Demosaicing Methods. Advances in Imaging and Electron Physics - ADV IMAG ELECTRON PHYS. 2010;

[18] Zhang L, Wu X, Buades A, li X. Color demosaicking by local directional interpolation and nonlocal adaptive thresholding. Journal of Electronic Imaging - J ELECTRON IMAGING. 2011 Jan 1;

[19] Ghimpeteanu G, Batard T, Levine S, Bertalmío M. Three Approaches to Improve Denoising Results that Do Not Involve Developing New Denoising Methods. In: Bertalmío M, editor. Denoising of Photographic Images and Video: Fundamentals, Open Challenges and New Trends. Springer International Publishing; 2018. p. 295–329.

[20] Buades A, Lisani J-L. Patch-Based Methods for Video Denoising: Fundamentals, Open Challenges and New Trends. In: Denoising of Photographic Images and Video: Fundamentals, Open Challenges and New Trends. 2018. p. 175–205.

[21] Zuo W, Zhang K, Zhang L. Convolutional Neural Networks for Image Denoising and Restoration: Fundamentals, Open Challenges and New Trends. In: Denoising of Photographic Images and Video: Fundamentals, Open Challenges and New Trends. 2018. p. 93–123.

[22] Buades A, Lisani J, Miladinović M. Patch-Based Video Denoising With Optical Flow Estimation. IEEE Transactions on Image Processing. 2016 Jun;

[23] Buades A, Coll B, Morel J-M. A Review of Image Denoising Algorithms, with a New One. SIAM Journal on Multiscale Modeling and Simulation. 2005 Jan 1;

[24] li X, Dong W, Shi G. Sparsity-Based Denoising of Photographic Images: From Model-Based to Data-Driven: Fundamentals, Open Challenges and New Trends. In: Denoising of Photographic Images and Video: Fundamentals, Open Challenges and New Trends. 2018. p. 37–62.

[25] Moeller M, Cremers D. Image Denoising—Old and New: Fundamentals, Open Challenges and New Trends. In: Denoising of Photographic Images and Video: Fundamentals, Open Challenges and New Trends. 2018. p. 63–91.

[26] Tan H, Zeng X, Lai S, Liu Y, Zhang M. Joint demosaicing and denoising of noisy bayer images with ADMM. 2017;

[27] Alkinani MH, El-Sakka MR. Patch-based models and algorithms for image denoising: a comparative review between patch-based images denoising methods for additive noise reduction. EURASIP Journal on Image and Video Processing. 2017 Dec 1;

[28] Dutra da Silva R, Minetto R, Schwartz W, Pedrini H. Adaptive edge-preserving image denoising using wavelet transforms. Pattern Analysis and Applications. 2013 Nov 1;

[29] Tsai RY, Huang TS. Multiframe image restoration and registration. Advances in computer vision and image processing [Internet]. 1984 [cited 2019 Apr 17]; Available from: https://ci.nii.ac.jp/naid/10027289457

[30] Tico M, Pulli K. Image enhancement method via blur and noisy image fusion. 2009 Dec 10;

[31] W. Hasinoff S, Sharlet D, Geiss R, Adams A, T. Barron J, Kainz F, et al. Burst photography for high dynamic range and low-light imaging on mobile cameras. ACM Transactions on Graphics. 2016 Nov 11;

[32] Boracchi G, Foi A. Multiframe Raw-Data Denoising Based on Block-Matching and 3-D Filtering for Low-Light Imaging and Stabilization. 2008.

[33] Night Sight: Seeing in the Dark on Pixel Phones [Internet]. Google AI Blog. [cited 2019 Apr 18]. Available from: http://ai.googleblog.com/2018/11/night-sight-seeing-in-dark-on-pixel.html

[34] Liu Z, Yuan L, Tang X, Uyttendaele M, Sun J. Fast Burst Images Denoising. ACM Transactions on Graphics. 2014 Nov 19;

[35] Khademi Kalantari N, Shechtman E, Barnes C, Darabi S, Goldman D, Sen P. Patch-Based High Dynamic Range Video. ACM Transactions on Graphics (TOG). 2013 Nov 1;

[36] Tico M. Multi-frame image denoising and stabilization. 2008 Jan 1;

[37] Liu S, Yuan L, Tan P, Sun J. Bundled Camera Paths for Video Stabilization. ACM Transactions on Graphics. 2013 Jul 1;

[38] Trimeche M. Super-resolution image reconstruction using non-linear filtering techniques. 2006.

[39] WolfWings. Pincushion distortion visual example [Internet]. 2008 [cited 2019 Apr 18]. Available from: https://commons.wikimedia.org/wiki/File:Pincushion_distortion.svg

[40] Lee JS, Hoppel K. Noise Modeling and Estimation of Remotely-Sensed Images. 12th Canadian Symposium on Remote Sensing Geoscience and Remote Sensing Symposium,. 1989 Jul;

[41] Gravel P, Beaudoin G, Guise JAD. A method for modeling noise in medical images. IEEE Transactions on Medical Imaging. 2004 Oct;

[42] Mäkitalo M, Foi A. Optimal Inversion of the Generalized Anscombe Transformation for Poisson-Gaussian Noise. IEEE transactions on image processing. 2012 Jun 5;

[43] R. Isidoro J, Milanfar P. Pull-Push Non-local Means with Guided and Burst Filtering Capabilities: Fundamentals, Open Challenges and New Trends. In: Denoising of Photographic Images and Video: Fundamentals, Open Challenges and New Trends. 2018. p. 267–94.

[44] Buades T, Lou Y, Morel JM, Tang Z. A note on multi-image denoising. 2009;

[45] Lowe D. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision. 2004 Nov 1;

[46] Vandewalle P, Süsstrunk S, Vetterli M. A Frequency Domain Approach to Registration of Aliased Images with Application to Super-resolution. EURASIP Journal on Advances in Signal Processing. 2006 Dec;

[47] Mildenhall B, T. Barron J, Chen J, Sharlet D, Ng R, Carroll R. Burst Denoising with Kernel Prediction Networks. 2017;

[48] Bovik AC, Books24x7 I. Handbook of Image and Video Processing, 2nd Edition. 2;2nd; Burlington, Mass: Academic Press; 2010.

[49] Google Camera Port: Arnova8G2 apks [Internet]. [cited 2019 Apr 18]. Available from: https://www.celsoazevedo.com/files/android/google-camera/dev-arnova8G2/

[50] HDR+: Low Light and High Dynamic Range photography in the Google Camera App [Internet]. Google AI Blog. [cited 2019 Apr 17]. Available from: http://ai.googleblog.com/2014/10/hdr-low-light-and-high-dynamic-range.html

[51] Dabov K, Foi A, Katkovnik V, Egiazarian K. Color Image Denoising via Sparse 3D Collaborative Filtering with Grouping Constraint in Luminance-Chrominance Space. 2007 IEEE International Conference on Image Processing. 2007;

[52] Lebrun M, Buades A, Morel J-M. Implementation of the ‘Non-Local Bayes’ (NL-Bayes) Image Denoising Algorithm. Image Processing On Line. 2013 Jun 17;

[53] Maggioni M, Boracchi G, Foi A, Egiazarian K. Video Denoising, Deblocking, and Enhancement Through Separable 4-D Nonlocal Spatiotemporal Transforms. IEEE Transactions on Image Processing. 2012 Sep;

[54] Image and video denoising by sparse 3D transform-domain collaborative filtering | Block-matching and 3D filtering (BM3D) algorithm and its extensions [Internet]. [cited 2019 Apr 18]. Available from: https://www.cs.tut.fi/~foi/GCF-BM3D/

[55] Video Denoising with Optical Flow Estimation [Internet]. [cited 2019 Apr 18]. Available from: http://www.ipol.im/pub/art/2018/224/

[56] Implementation of the ‘Non-Local Bayes’ (NL-Bayes) Image Denoising Algorithm [Internet]. [cited 2019 Apr 18]. Available from: https://www.ipol.im/pub/art/2013/16/

[57] Middlebury dataset [Internet]. [cited 2019 Apr 18]. Available from: http://vision.middlebury.edu/flow/data/

[58] Plötz T, Roth S. Benchmarking Denoising Algorithms with Real Photographs. 2017 Jul 5;